Available online at www.sciencedirect.com
SCIENCE DIRECT®
Field Crops Research 92 (2005) 119–135
Field Crops Research
www.elsevier.com/locate/fcr

Integration of sucrose accumulation processes across hierarchical scales: towards developing an understanding of the gene-to-crop continuum

Paul H. Moore

Plant Physiologist, USDA, ARS, Pacific Basin Agricultural Research Center, Aiea, HI, USA

Abstract

Although we have a working knowledge of the general biology of sugarcane and are fairly efficient in crop production, we are only beginning to produce the detailed biochemical and genetic information that will be needed to develop technologies necessary for long-term success of the sugarcane industry. A true understanding of sucrose accumulation will be achieved only through a combination of experimental and computational analyses of comprehensive datasets to gain information on the underlying molecules and processes. Currently, systems-level approaches towards understanding biology are gaining momentum primarily because of the tremendous increases in computing power, advances in information sciences and rapid progress being made in molecular biology to facilitate data management, analyses and modeling from high-throughput genome sequencing and proteomics. Large-scale sequencing projects have not only provided complete sequence information for a number of genomes, but they are also producing integrated pathway-genome databases and models that provide organism-specific connectivity maps of metabolic and, to a lesser extent, other cellular networks. To date, the processes that generate mass, energy, information transfer and cell-fate specification have been analyzed only at the cell or microorganism levels, where they are shown to be seamlessly integrated through a complex network of cellular constituents and reactions. Despite the key role of these networks in sustaining cellular functions, their large-scale structure is essentially unknown. Fifty years of sugarcane research have identified and characterized a suite of physiological processes and enzymes involved in sucrose accumulation. More recently, genes encoding these enzymes have been isolated, cloned and used in experiments to transform sugarcane to increase or decrease expression of the enzyme with the goal of altering sucrose accumulation. However, results of this reductionist approach towards understanding sucrose accumulation have fallen short of expectations, apparently because of the complex interactions among the multitude of simultaneous processes. Recent rapid expansion of sugarcane molecular datasets and the beginning of a systems approach to metabolic modeling of sucrose accumulation point the way for future research efforts to integrate processes from gene to crop performance.

Keywords: Systems biology; Sugarcane; Saccharum; Yield improvement; Molecular databases

0378-4290/$ – see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.fcr.2005.01.031

1. Introduction

''Forget about the genome. And let's say, 'It's great. We've got all these sequences. Thank you very much for all the people that helped to get them—please get us some more.' And let's get on with the biology.'' Sydney Brenner, speaking at the Plant and Animal Genome Conference, San Diego, January 2003.

There was great fanfare in 2003, and deservedly so, in recognizing the 50th anniversary of the discovery of the structure of DNA (Watson and Crick, 1953). This intellectual achievement has become the basis for the concepts and technologies of molecular biology that link information about life's physical entities such as DNA or proteins to the genetic traits of organisms. Advances in molecular approaches have revolutionized biology to such a degree that a report from a January 19, 2003, U.S. Department of Energy workshop on biosciences states that more has been learned this past decade about how plants work than in all of preceding history (Minorsky, 2003). Not only does this statement ring true, I believe that advancements made during the next decade will tower over those of the last. The specificity, speed and minuteness of scale of molecular biology research have led to development of high-throughput technologies that quickly provide us spectacular breakthroughs such as identifying all the genes in a plant. This was demonstrated recently by the full sequencing of two plant genomes—the model plant Arabidopsis (*Arabidopsis thaliana*) and the crop plant rice (*Oriza sativa*). In addition, post-genomic developments, such as using microarrays and proteomic techniques, are providing us the ability to determine which genes are activated or inactivated during development or in response to an environmental change.

Amassing large molecular datasets is one thing; producing an understanding of what it all means about plant growth and development is quite another. Collecting and analyzing the information from even a single high-throughput experiment quickly reveals the complexity and magnitude of effort required to build a sufficiently comprehensive dataset to predict how the plant, or tissue, or cell, or biochemical pathway will perform under every different set of conditions. One can scarcely envision the magnitude of systematically establishing perhaps 500 different environmental or developmental stages over which to compare the mRNA expression of a genome of 25,000 genes. Explaining a biological system by knowing details of its many components has been compared to trying to build a computer or a Boeing 777 by having blueprint specifications only for each of the components. What is needed in place of a traditional bottom-up approach, i.e., analyzing the details of small subsystems, is a top–down approach of modeling the entire system from the behavior of its many subsystems. In molecular biology, a top–down or ''systems approach'' is based on models that are developed to integrate the exhaustive descriptions of biological systems to predict how the different layers or hierarchical scales interact to form higher functional units like coordinated pathways, regulatory networks or complex structures such as cells or tissues. The ultimate goal is to understand the biological system in sufficient detail to enable accurate, quantitative predictions about its behavior when we somehow manage to introduce or block the expression of a suite of genes. This will give us the ability to engineer the design of a crop plant predictably. The challenge is to improve our understanding of how plants function at all scales of complexity to such an extent that we can produce models that will predict how a crop will respond to any given genetic manipulation or environmental perturbation.

The need to conduct research to engineer the design of crops for increased productivity is related to the world's predicted population growth, its finite environmental resources and the apparent inadequacy of current breeding and selection technologies for developing crop varieties with higher yields, on more marginal lands, and having reduced consumption of water and nutrients. The world's population is currently more than 6.4 billion and is projected to exceed 7.8 billion during the next decade. The already inadequate production of food and fiber to meet population needs is becoming more acute with the increasing demands of a rapidly increasing population. Fortunately, just at the time we appear to be testing the limits of crop improvement by conventional approaches, new approaches are being offered by the emergence of high-throughput molecular technologies that are rapidly expanding our knowledge of biological systems and suggest ways to control them for greater production.

The purpose of this paper is to review briefly the history of crop yield improvements and to relate other crop yields to those of sugarcane to highlight the pressing need to develop alternatives to past methods for increasing crop yield potentials. I will explore recent plant molecular biology research developments that offer clues about what these alternative methods might be, outline some of the missing elements that need to be developed and suggest approaches that might be explored by the sugarcane research community to increase yields of the sugarcane crop.

2. Crop yields

Before outlining how we might improve crop productivity, it is worthwhile to step back to look at where we are with crop yields, how we got here and where the opportunities lie for making improvements. Crop yield statistics are a place to begin, even though we recognize that these statistics do not address what yields are achievable since they are only a grand mean of actually reported yields over a wide range of conditions. Therefore, we need separate data from a range of crop production systems (Rabbinage, 1993) to identify yield constraining factors that might be managed to increase yields. Actual yields are those reached under the yield-limiting constraints of various pests, diseases and soil nutrient deficiencies for which ameliorations are generally available (Fig. 1). Relatively small inputs of fertilizers and pesticides can raise actual yields to approach yields attainable under the prevailing environment. Attainable yields are constrained by the prevailing environment that might include suboptimal factors such as limitations of water, radiation, temperature and soil inhibitory factors such as mineral toxicity or salinity/sodicity. Ameliorating the environmental constraints to raise attainable yields to approach the potential yield is more difficult and costly than raising the actual yield, but it is commonly done in advanced agronomic situations. Potential yield is the yield that is achieved when the crop is grown with an ample supply of water and nutrients and the absence of pests; one can

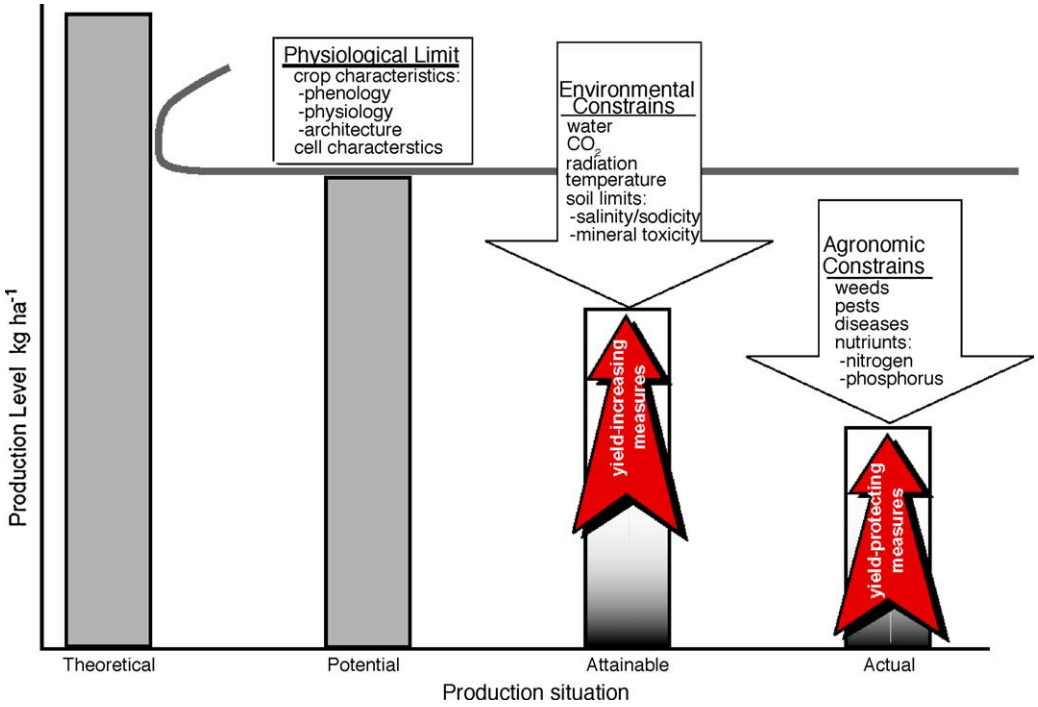

Fig. 1. Crop production situations, production levels, associated principal factors limiting production and agronomic inputs with potential for protecting or increasing yields (modified from Rabbinage, 1993).

consider record yields as approaching or the equivalent of potential yields. However, even this level of yield is lower than the theoretical yield maxima that have been calculated from models of physiological processes contributing to plant growth (Loomis and Amthor, 1999). The relationship among these production situations can be used to identify where R&D resources might give the greatest return in increasing crop yields. Under poor production situations, yields will likely be increased by minimizing the effects of reducing factors such as pests and diseases, and then satisfying the limiting factors such as water and nutrients. Under advanced production systems, those external reducing and limiting factors may already be addressed so that greatest return might be from R&D aimed at altering the genetics of the crop plant to raise the potential yield.

Although we generally speak of plant breeding as a science-based activity dating from the early 20th century and the rediscovery of Mendel's principles, in reality plant selection-breeding for crop improvement was a prerequisite for civilization. For at least 10,000 years, mankind has molded crop phenotypes by crossing among the best varieties and selecting those best adapted for dependable production under local conditions. Although long-term crop yield records are rare, they have been reported for 12 centuries of rice production in Japan and for 8 centuries of wheat production in England (Fig. 2a). From these widely distributed data, it is obvious that the rate of crop yield improvement was quite slow before undergoing a very rapid increase around the 1900s. Likewise, a plot of only the last 100 years for maize shows greater than four-fold yield increases during the first half of the 1900s (Fig. 2b). Record yields of maize in the United States remained essentially unchanged, at below 1.5 t ha^{-1}, until the 1930s. The corn yield ceiling was broken with the application of the principles of genetics and development of hybrid lines (Duvick, 1996). Likewise, with sugarcane the breakthrough in yields came about with the hybridization between *Saccharum officinarum* and *Saccharum spontaneum* at the turn of the 20th century. Since then there has been a two- to three-fold increase in yields. Similar yield increases were common with other crops in the United States, Western Europe and Japan. However, the increases did not occur in the less developed countries until the early 1970s with Norman Borlaug's movement called the ''Green Revolution'' based on dwarf hybrid cereal crops, primarily rice, capable of using higher inputs of inorganic nitrogen and partitioning the increased productivity to seed instead of to straw.

Notwithstanding the great advances in crop yields arising from the technologies of the Green Revolution, there is now conflicting evidence whether crop-yield potential has increased in recent years (Austin, 1999; Duvick and Cassman, 1999; Evans and Fischer, 1999; Peng et al., 1999; Reynolds et al., 1999; Specht et al., 1999; Tollenaar and Wu, 1999). Maximum yield trials for rice at the International Rice Research Institute in the Philippines have not risen since the early 1980s (Pingali et al., 1990). Grain sorghum yields increased 11% from 1950 to 1960; 4% from 1961 to 1970 and only 2% from 1971 to 1980 (Miller and Kebede, 1984). An even more dramatic decline in crop yield improvement is reported for U.S. cotton lint. Yields increased 10.4 kg ha^{-1} $year^{-1}$ from 1936 through 1960 but declined an average of 0.92 kg ha^{-1} $year^{-1}$ from 1961 through 1980 (Meredith and Bridge, 1984). For maize in the United States, any yield increase would be a much greater percent in 1930, when the base yield was only 1.5 t ha^{-1}, than it is today when the base is 8.5 t ha^{-1} (Fig. 2b). Sugarcane may also be approaching a yield plateau in the more productive regions such as Hawaii. Australia has recently achieved the production levels of Hawaii. The ability to sustain these yields will be a test of whether these levels represent some kind of yield ceiling on sugarcane (Fig. 3).

The general crop science literature indicates that crop yields obtained by the better farmers in favorable areas or at experiment station maximum yield trials, are not far below a physiological limit (Sinclair and Muchow, 1999). If the present yield ceilings are to be broken, it seems unlikely to be achieved through additional management and technical inputs such as fertilizer, pesticides and irrigation. Nor is it likely that future yield increases will come from traditional genetic improvements such as optimizing plant architecture to accommodate higher nitrogen applications to short stature plants or designing plants with erect leaves that enable higher plant populations per unit area or from incorporating major genes for resistance to pest and pathogens. Instead, future advances will likely be from areas previously

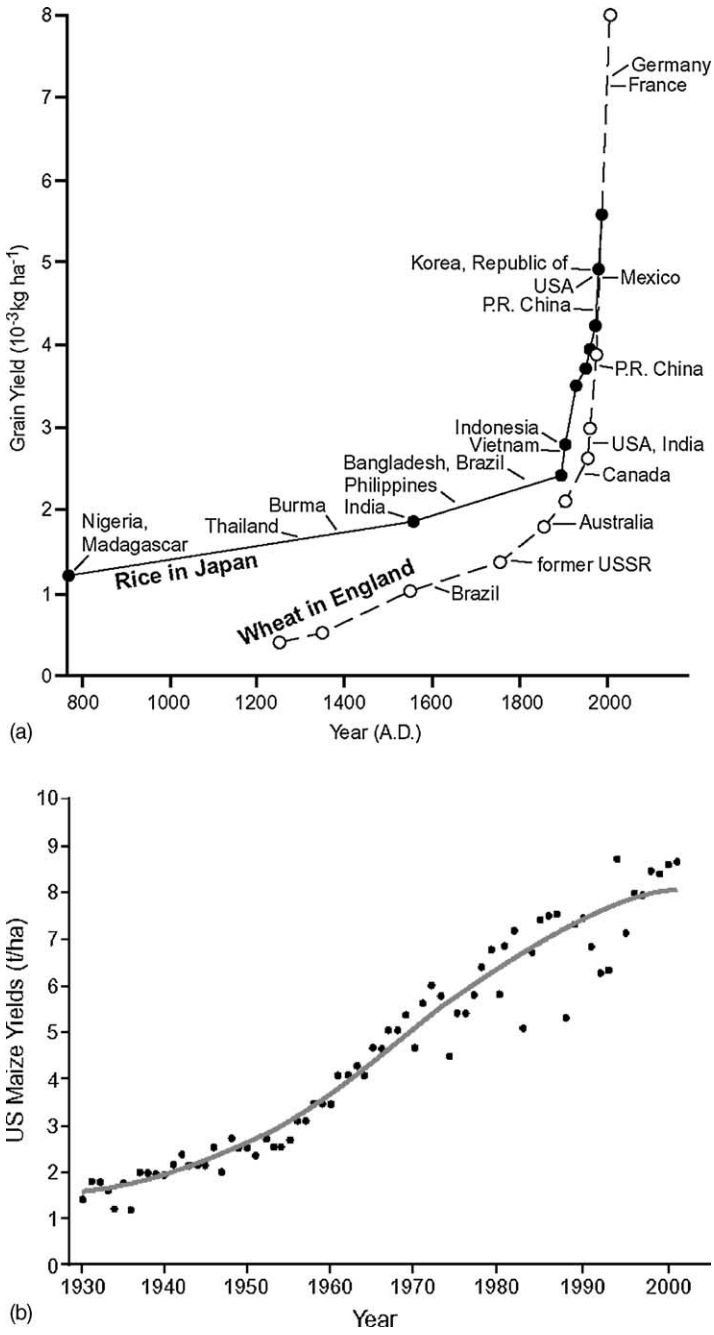

Fig. 2. Historical yield data for specific crops in various countries. (a) Historical yields of rice in Japan from 800 to 1950 A.D. and wheat in England from 1200 to 1950 (modified from Evans, 1975). The 2000/2001 average yields of these crops in various countries are superimposed along the historical yield curves (2000/2001 yield data from USDA-NASS, Agricultural Statistics Handbooks, 2004, <http://www.usda.gov/nass/pubs/agr04/acro04.htm>). (b) Three-year simple moving average of yields of U.S. maize fitted to a smoothing function (data from USDA-NASS Agricultural Statistics Handbooks, 2004, <http://www.usda.gov/nass/pubs/agr04/acro04.htm>).

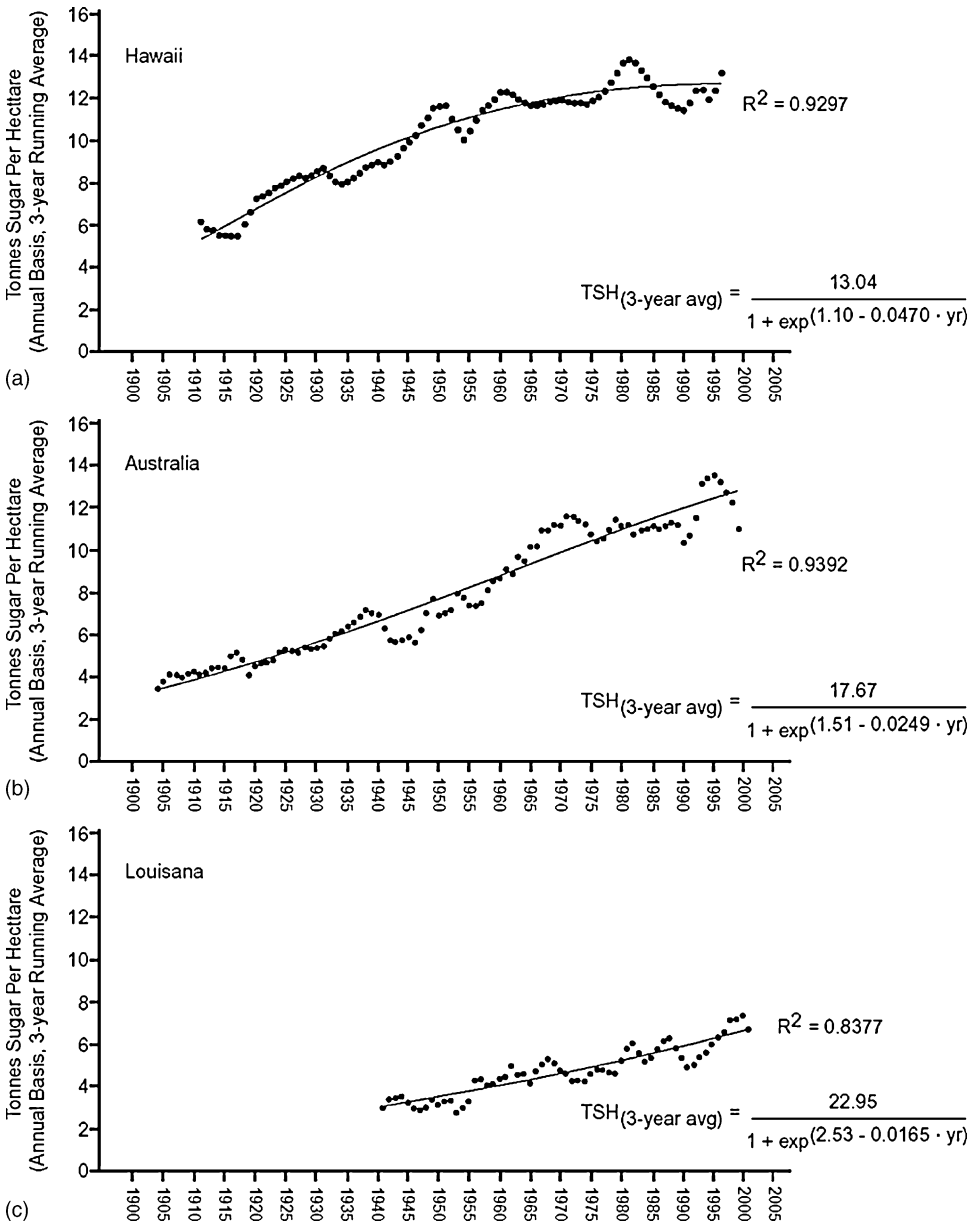

Fig. 3. Historical yield data for cane sugar production in (a) Hawaii (data from Hawaiian Sugar Manual 1995, Hawaiian Sugar Planters' Association), (b) Australia (data from Cox et al., 2000) and (c) Louisiana (data from Tom Tew, USDA, ARS, Sugarcane Station, Houma, LA, ttew@srrc.ars.usda.gov). Data points are simple 3-year running averages. The best-fit smoothed curves were calculated from equations shown.

unattainable because of their genetic complexity. Recent advancements in high throughput analyses leading to the establishment of large databases and new methods for analyses have partially resolved these complexities in some model prokaryotes and show potential for application to crop plants. For example, we might see genetic improvements in photosynthetic efficiency, partitioning of photosynthates among metabolic pools and quantitative resistance to pests or pathogens.

3. Sugarcane yields

Improvement of sugarcane yields through breeding and selection has been a directed, ongoing process following the observation in 1858 that sugarcane panicles produce viable seed. In 1888, the Dutch established an innovative breeding and selection program in Java to incorporate the disease resistance, hardiness and tillering capacity of *S. spontaneum* into the sugar producing germplasm of *S. officinarum*. The key event of this effort was the release in 1921 of the first of the nobilized hybrid cane cultivars, POJ 2725 and POJ 2878. These two early cultivars have served as the foundation in the pedigree of nearly all locally developed and adapted modern sugarcane cultivars world-wide.

The spectacular early successes of the newly developed hybrids encouraged most sugarcane-producing countries to establish sugarcane experiment stations to produce locally adapted cultivars and to do research to develop cultural practices for optimizing production. The combination of new, locally adapted varieties and improved agronomic practices resulted in significant increases in cane and sugar yield, disease and insect resistance, stress tolerance and other traits contributing to yield (Fig. 3). A jump in yields may be followed by a gradual slowing of the sort that has been in place in Hawaii for perhaps 20 years. One explanation for the apparent slowing of yield increases in Hawaii is that agronomic inputs such as fertilizers and weed control have been decreased on lands being taken out of sugarcane production by a shrinking industry. An alternative explanation is that the yields attained in Hawaii are such a large fraction of the physiological yield possible that the percentage gain of each generation of crop improvement is small. In other words, yields are approaching a yield ceiling. If this is so, then we might expect a leveling out of the Australia yields that have recently reached the levels of those in Hawaii.

4. Hierarchical scaling and metabolic system networks

Because plants are central to human survival, people must always have marveled at and sought explanations for the shifting patterns of plant development across seasons and environments. The gradual formalization of such inquiry into the scientific method led to explanations about plant development on the bases of anatomy and morphology, cell biology, biochemistry, physiology, genetics and molecular biology. As the analyses of plant processes produced ever more detailed knowledge, each field of science developed its own language, concepts and principles to explain plant processes at a particular magnitude of scale. The scalar relationship of each field of science to all others, organizes into a linear hierarchy (Fig. 4) that aims to integrate fundamental facts verified at lower levels into hypotheses about behavior at the higher levels. On the other side of the coin, there is the simultaneous need to explain the behavior of the higher level by summarizing or reducing its apparent complexity to the more fundamental behavior of the lower levels. Thus, the concept of ‘scaling’ is an attempt to integrate and consolidate knowledge across scales of organizational complexity. As mentioned earlier in the analogy about how a complete list of the parts is not sufficient to understand how a jumbo jet airplane functions, a detailed molecular knowledge about a plant process is not sufficient to predict behavior at the tissue, organ, plant or crop levels. Obviously, large-scale organizational levels have characteristics that cannot be predicted from knowledge of the small scale. While scaling beyond one or two levels is theoretically possible, it is achievable only through experimentation that may be so extensive as to be impractical. An alternative for applying knowledge at the gene or metabolite level to processes at the crop level would be to employ a systems approach that is scale-neutral.The central dogma of molecular biology is that information stored in DNA is processed through RNA to produce proteins that execute various cellular programs. At the hierarchical level the information flow is from the genes, to mRNA, proteins and eventually the metabolites (Fig. 5). Recently, ideas about linear flow of information have been revised to recognize that although long-term information storage may be almost exclusively in the genome, information for regulation of gene expression frequently flows back to the genome from the transcriptome, proteome and metabolome. This understanding is forcing biologists to develop a more integrated view of cellular functions as being distributed among groups of non-identical elements

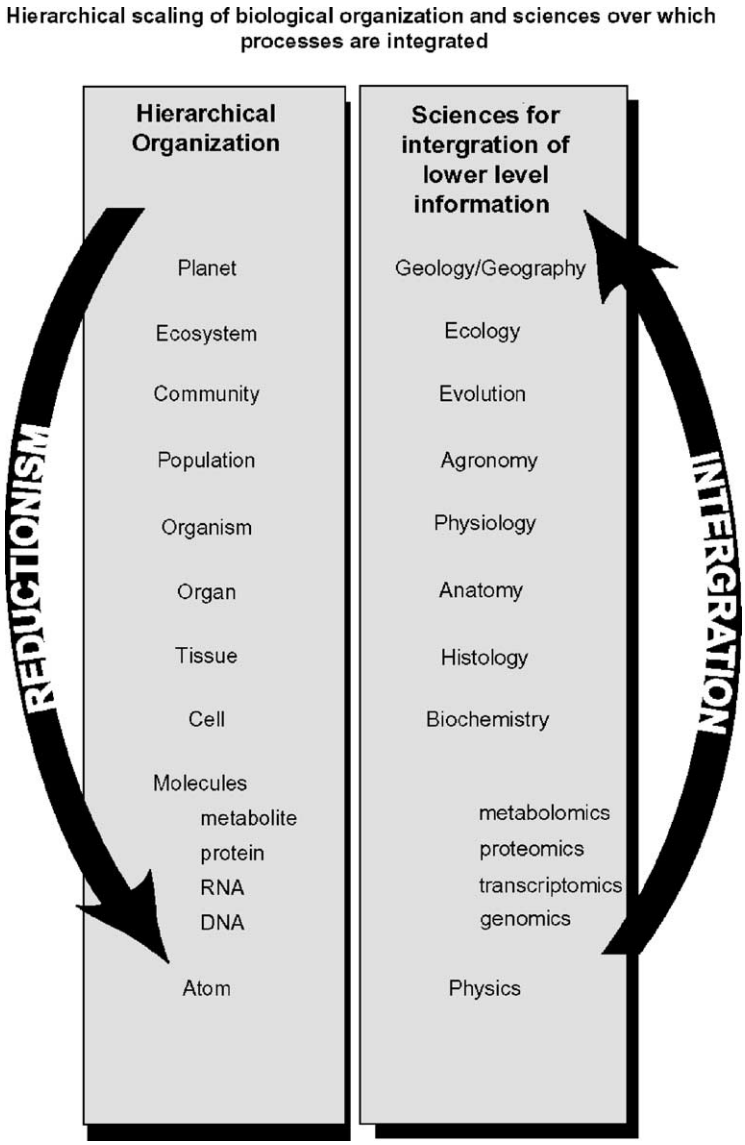

Fig. 4. Hierarchical scaling of biological organization and the sciences over which processes are integrated through gene expression.

that all interact within large networks (Oltvai and Barabasi, 2002). The thousands of components of a living cell are dynamically interconnected, so that the cell's functional properties exist as a complex intracellular web of molecular interactions. This is most evident when viewing a wall chart of cellular metabolism where hundreds of metabolic substrates are densely interconnected in pathways of biochemical reactions. What cannot be shown in these metabolic charts, but should be obvious on reflection, is that because a specific substrate can react with another substrate and specific proteins interact in many biochemical reactions, metabolic networks are not random but are ordered in a specific topology that can be organized into subsections referred to as pathways involving enzymes, substrates and products under regulation by a network motif of genes (Milo et al., 2002; Ravasz et al., 2002).

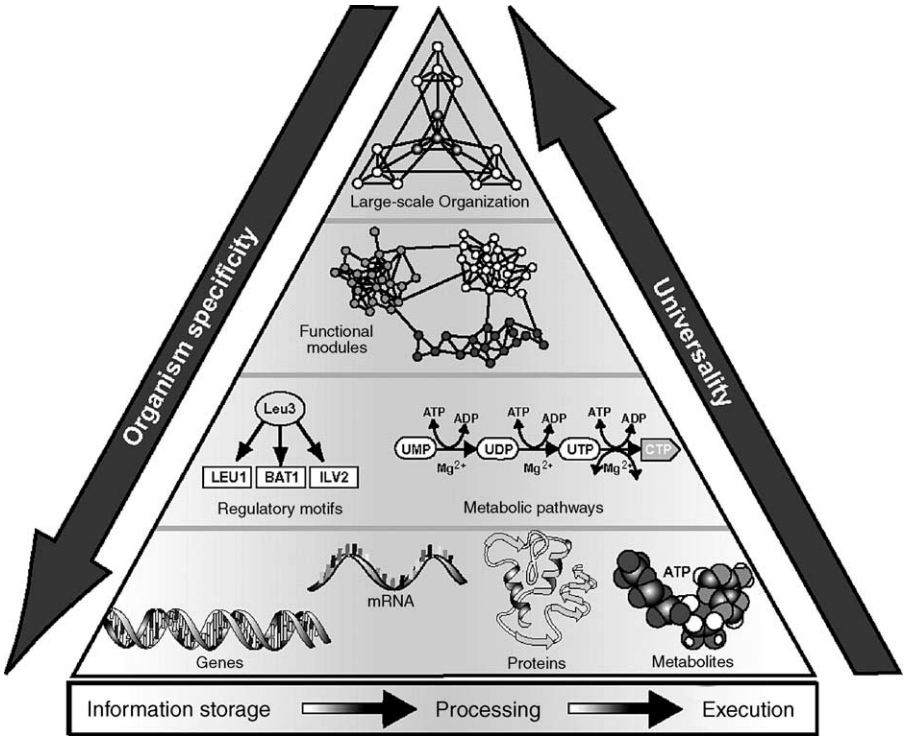

Fig. 5. Life's complexity pyramid from the particular to large-scale organization. Base of the pyramid is the fundamental flow of information from the genes, to mRNA transcripts, proteins and metabolites that are coordinated at the next higher level by regulatory motifs and metabolic pathways that are linked at the next higher level into functional modules that in turn are assembled into large-scale organization. Although the individual components are unique to a given organism at a particular time and stage of development, the topological properties share similarities with other large-scale processes. Reprinted with permission from Oltvai and Barabasi (2002). © 2002 American Association for the Advancement of Science. Credit K. Sutliff.

When the pattern of interconnections of elements (called nodes) in a complex network occurs at numbers that are significantly higher than those in randomized networks, it is called a network motif. If the interconnections of nodes were random, then each pair of nodes would have the same probability p of connecting and would follow a Poisson distribution (Fig. 6). This is not what is seen in empirical studies of large-scale networks as diverse as the worldwide web, food webs, social interactions and metabolic networks (Milo et al., 2002). In these large-scale networks, the probabilities of interconnections among nodes decay exponentially according to a power law showing that most nodes have only a few links, but a few of them have a very large number of links. Although the concept of networks and the assembly of nodes into motifs is not generally recognized by the public, this aspect can be enjoyed by them as the basis of a popular diversion called the ''Kevin Bacon Game'' http://www.louisville.com/loumag/mar/bacon.htm or less specifically, ''six degrees of separation.'' According to the rules of this game, no more than six connections should be sufficient to describe the interconnections of very large-scale networks, such as the human population of the world. This result is a demonstration that large networks self-organize into a scale-free state, a feature unpredicted by all existing random network models (Barabási and Albert, 1999). This recently discovered insight is now being used to develop new models based on information of genomics and proteomics to interconnect biological elements and processes and across hierarchical scales into testable models that can be used to predict how cells and organisms will behave when perturbed.

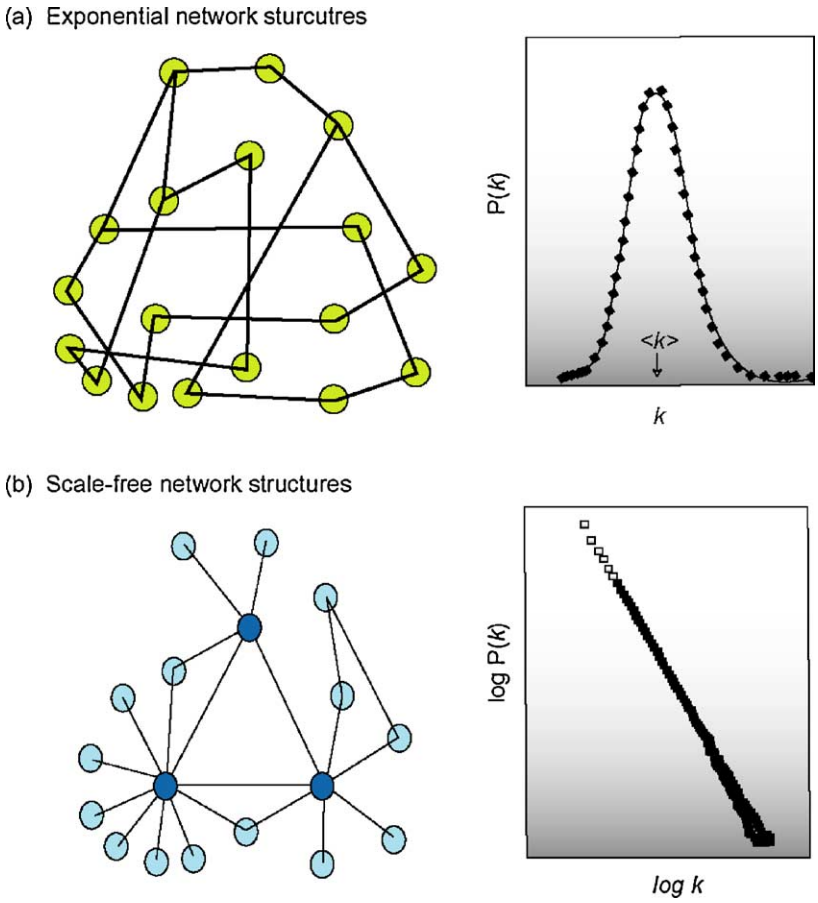

Fig. 6. Contrasting attributes of (a) exponential vs. (b) scale-free generic network structures. Network connectivity can be characterized by the probability, $P(k)$, that a node has k links. For an exponential network, $P(k)$ peaks strongly at $k = <k>$ and decays exponential for large k. In the scale-free network, most nodes have only a few links, but a few nodes, called hubs, each of which have a very large number of links. In this case, $P(k)$, has no well-defined peak and for large k it decays as a power-law appearing as a straight line on a log–log plot. Reprinted with permission from Jeong et al. (2000). http://www.nature.com.

The biological generality that metabolic networks self-organize was convincingly shown by Jeong et al. (2000) in a study of the core metabolic network of 43 different organisms based on the intermediate metabolism and bioenergetics data deposited in the WIT database (Overbeek et al., 2000). Jeong showed that the various molecular components of the cell–genes, RNAs, proteins and metabolites—organize themselves into small recurrent patterns called pathways in metabolism or motifs in genetic-regulatory networks. In turn, the motifs and pathways are seamlessly integrated to form functional modules consisting of proteins and metabolites that are responsible for discrete cellular functions. These modules are nested hierarchically to define the cell's large-scale functional organization (Jeong et al., 2000). This may indicate that metabolic organization is not only identical for most organisms, but that it also complies with the design principles of robust and error-tolerant, scale-free networks that can serve as a model for the large-scale organization of interactions among all cellular constituents.

Approaches to integrate large datasets such as those from DNA microarrays and quantitative proteomics were developed to build, test and refine a model of

cellular pathways in yeast (Ideker et al., 2001). The general procedure consisted of these four steps:

- Accumulate information on all of the genes of the genome and the subset of genes, proteins and other small molecules that might be active in the pathway of interest. Use these data and knowledge of previous research to develop an initial model of the molecular interactions involved in the pathway.
- Perturb a number of the pathway components either endogenously by producing genetic changes (gene deletions, gene silencing or gene overexpression) or exogenously by altering the environment (temperature, nutrient supply, pH or light). Then use high throughput technologies such as large-scale expression of mRNA or proteins to detect and quantify the global cellular response to each perturbation.
- Integrate the observed mRNA and protein responses with the current pathway model and with knowledge of protein–protein, protein–DNA and other known physical interactions.
- Revise the initial hypothesis to explain observations not predicted by the model. Design additional perturbation experiments to test the new hypothesis and repeat each step until a stable network of interactions can be developed (Fig. 7).

Using this approach, Ideker et al. (2001) identified 997 mRNAs responding to 20 systematic perturbations of the yeast galactose-utilization pathway (Fig. 8). They detected involvement of 289 proteins of which 15 were regulated posttranscriptionally. This information was used to refine the initial model to develop a global model of galactose utilization and physical interactions between this pathway and a variety of other pathways.

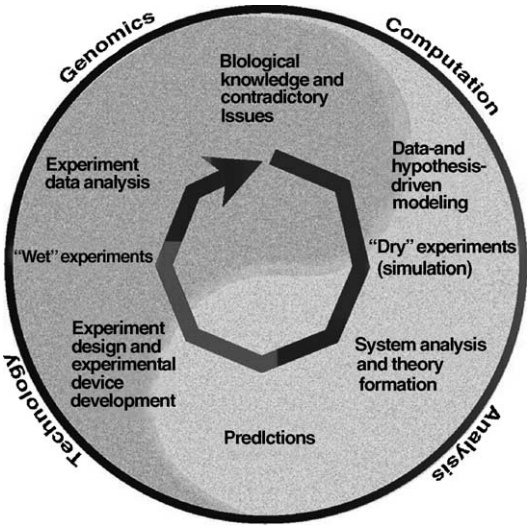

Fig. 7. The cycle of systems biology involving hypothesis, model building and experimental verification for increased understanding of complex biological systems. The model represents a computable set of assumptions and hypotheses that need to be tested or supported experimentally. Computational ''dry'' experiments such model simulations reveal computational adequacy of the assumptions and hypotheses embedded in each model. Inadequate models expose inconsistencies with experimental facts and need to be rejected or modified. Models that pass this test are selected for ''wet'' experiments that produce additional data to support or reject the evolving hypotheses that are iteratively refined to develop biological knowledge. Reprinted with permission from Kitano (2002). © 2002 American Association for the Advancement of Science.

5. Sucrose accumulation processes

Sucrose synthesis and accumulation in higher plants is the product of a very large network of interactions that can be analyzed from several perspectives. At the gross level, sucrose accumulation is simply the difference between the amount of sucrose produced in the leaf by photosynthesis and the amount of this sucrose that is removed by metabolism to produce carbon and energy for growth and other components of the plant. But this is not all that sucrose does. There is increasing evidence that sucrose is involved in signaling to modulate expression of genes controlling cell division and differentiation, transporters and storage proteins, induction of flowering, differentiation of vascular tissue, seed development and accumulation of storage products (Lunn and MacRae, 2003). Each of the reactions involved is controlled by activation of specific genes by an interaction among the genotype of the plant, the environment under which it is growing and its developmental stage at that instant.

Modern sugarcane cultivars are interspecific hybrids that, under ideal conditions, are capable of storing sucrose in the parenchyma tissues of the stem up to 62% of the dry weight or 25% of the fresh weight (Bull and Glasziou, 1963). These are approximately the levels obtained in *S. officinarum*, the major source of commercial hybrid germplasm. On the other hand,

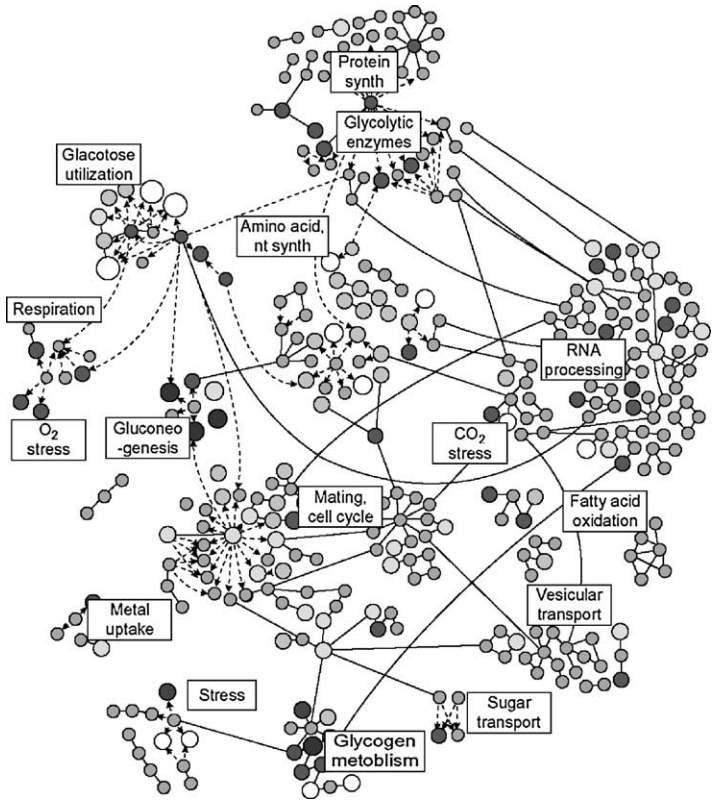

Fig. 8. Integrated genomic and proteomic physical-interaction network for yeast galactose (GAL) utilization. Nodes represent genes. A dashed arrow directed from one node to another signifies that the protein encoded by the first gene can influence the transcription of the second by DNA binding (protein → DNA) and the solid line between two nodes signifies that the corresponding proteins can physically interact (protein–protein). The grey scale intensity of the nodes represent changes in mRNA when GAL utilization was perturbed by altering the environment (temperature, nutrient supply, pH or light) in different genetic strains having either gene deletions, gene silencing or gene over expression; medium grey represents no change in mRNA, whereas darker shades represent increased expression and lighter shades represent reduced expression. Node diameter scales with the magnitude of change. Highly interconnected groups of genes appear as functional modules having common biological function as labeled. Reprinted with permission from Ideker et al. (2001). © 2001 American Association for the Advancement of Science.

some of the wild relatives of sugarcane store less than 2% of the fresh weight as sucrose. These striking differences in sink accumulation and storage activity cannot be explained by differences in photosynthetic rates in the leaves. The photosynthetic rates on an area basis of *S. spontaneum* have been reported as nearly twice those of *S. officinarum* and 30% greater than that of hybrid cultivars (Irvine, 1975). Since photosynthetic rates cannot explain the differences in sucrose storage, it might be explained by differences occurring within the stem or within the transport system(s) between the source of photoassimilates in the leaves and the deposition of those photoassimilates in the stalk sink.

Multiple pathways involved in sucrose accumulation in which there might be rate-limiting physiobiochemical reactions include:

- leaf reactions including photosynthetic rate, sucrose synthesis, metabolism and carbon partitioning across various membranes into different pools;
- phloem reactions including loading in the leaf, translocation to and unloading in various sink tissues including primary storage in parenchyma cells of the stalk;
- stalk reactions including membrane transport, sucrose metabolism, carbon partitioning into different pools and remobilization of stored sucrose;

- genetic and developmental controls including timing of maturation;
- environment perception and signal transduction pathways to coordinate plant development.

The sucrose accumulation pathway involves metabolic and physical processes in cells and tissues that are involved in sucrose synthesis in the leaf, short-distance and long-distance transport tissues involved in the export of sucrose from the leaf to various sites in the plant and parenchyma tissues of the stem storage or meristem tissues located primarily at the shoot and root apices. Sucrose accumulation rates vary over a wide range as a function of plant genotype, developmental stage and the environment in which the plant is grown. The plant perceives and responds to its environment through a complex network for the perception and transduction of information to coordinate plant development including the accumulation of sucrose.

The size and complexity of the numerous networks involved in regulating sucrose accumulation have until very recently forced scientists to confine their studies to only a small subset of conditions presumed to be critical to sucrose accumulation. Sugarcane literature abounds with reports of the level of sucrose determined at the crop level or in plants or parts of plants as a function of various factors such as genotype and age of the plant and sometimes a set of differing environmental conditions such as different levels of mineral nutrients (frequently nitrogen), amounts of water and temperature. Some studies include reports correlating the activity of a small set of leaf or stem enzymes to the level of sucrose measured. There is also considerable literature evaluating enzyme activity and cell membrane properties based on studies with stem tissue slices, cell cultures and protoplasts isolated from cell cultures. Results of the numerous studies have produced a few simple models aimed at identifying bottlenecks or constraints on sucrose accumulation of sugarcane. Although these studies have produced considerable data and interesting hypotheses about sucrose accumulation, they explain only how a certain genotype under a prescribed set of conditions once behaved. While the data obtained thus far have value for establishing an initial hypothesis about sucrose accumulation, what is needed is a global analysis of the metabolic networks involved in sucrose accumulation. As outlined in the earlier section, the technologies for this are being rapidly developed and applied to other biological problems. In the section that follows it will be obvious that the data needed for a global analysis of sucrose accumulation are also being rapidly developed. What remains is to continue to build sugarcane databases and to incorporate the new sciences of systems biology and bioinformatics into the sugarcane research community.

6. Progress in sugarcane molecular biology

The application of DNA markers to genetic mapping for crop improvement began in the early 1980s. For sugarcane, the groundbreaking theoretical concept that single dose markers could be used for mapping any polyploid without knowing either the type or level of polyploidy (Wu et al., 1992) opened a new era for sugarcane genetic and genomic research. Now, just a little over a decade later, abundant genomic resources have been established, the basic chromosome numbers of *Saccharum* have been resolved, knowledge about the genetic diversity and structure of the sugarcane genome is significantly advanced and early generations of linkage maps have been developed. As the linkage maps developed, they revealed conserved genes and co-linearity of gene order with genomes of other grasses.

Co-linearity has been used to evaluate the correspondence of quantitative trait loci, QTLs, affecting related traits in sugarcane and other grasses. Sorghum, the closest relative of sugarcane, has a small diploid genome that has proven quite useful for sugarcane genome analysis. Corresponding QTLs controlling plant height and flowering were found in sorghum and sugarcane (Ming et al., 2002). Several previously mapped maize and rice mutations, and QTLs of the sugar metabolic pathway might be candidate genes for controlling sugar content in sugarcane (Ming et al., 2001). Sorghum, rice and maize linkage and physical maps were used to identify potential markers for fine mapping and chromosome walking towards cloning the rust resistance gene in sugarcane (Asnaghi et al., 2000, 2004); sorghum and rice synteny played a key role in developing a sugarcane BAC contig covering the resistance gene (Le Cunff et al., 2004). The close relationship among these grasses, a high degree of co-linearity and cross-hybridization of DNA probes are

compelling reasons for using the more abundant information from the small genomes of rice and especially sorghum to guide molecular mapping and positional cloning in sugarcane.

Current large-scale genome projects on a variety of plants, animals and microbes are making available vast amounts of information in the form of genomic sequence and expressed sequence tags (ESTs). National and international efforts underway to develop and catalog ESTs for major food crops such as rice, maize and soybean are paralleled by independent sugarcane EST projects in Australia, Brazil, South Africa and the United States. Most notable among them, accounting for >90% of reported sugarcane ESTs, is the SUCEST program of the Brazil ONSA consortium. An initial report about this herculean effort was published as a special issue of *Genetics and Molecular Biology* (Vol. 24, No. 1–4, 2001) entitled "Sugarcane Transcriptome: A Landmark in Plant Genomics in the Tropics." It contains 34 research articles from 74 sequencing and data mining laboratories relating sugarcane ESTs to factors such as flowering, signal transduction, plant development, aluminum toxicity, pest and pathogen defense systems, mitochondria and chloroplast functions, membrane transport and secretion, cell wall metabolism and cytoplasm metabolism. These analyses were based on the SUCEST database containing 238,000 ESTs from 26 sugarcane cDNA libraries constructed from several tissues—shoot apical meristem, flowers, lateral vegetative buds, unfurled immature leaves, mature leaves, roots, stem (culm) rind, stalk internodes, seed and tissue culture calli—at different developmental stages (Vettore et al., 2001). The ESTs similar in sequence were assembled into 43,000 clusters of which 38% had no matches in existing public sequence databases. Around 53% of the clusters were formed by ESTs in more than one library, delimiting a group of genes that are coordinately expressed in different tissues, while 47% were formed by ESTs expressed in only one library, delimiting tissue-specific expressed genes. A more thorough analysis of the SUCEST database (Vettore et al., 2003) reassembled the 43,000 sequences to report a 22% redundancy and tentative identity of 33,620 unique genes. Annotation of the 43,000 assembled sequences showed almost 50% of them were associated with protein metabolism, cellular communication/signal transduction, bioenergetics and stress response.

Because cane sugar is a major, internationally traded commodity, there is often a substantial lag in release of information that might lead to competitive advantages. The GenBank database for sugarcane currently lists more than 250,000 EST sequences with 4850 UniGenes. Interest in the possibility of discovering agronomically important sugarcane genes led an international consortium of sugarcane research institutions (the International Consortium of Sugarcane Biotechnology, ICSB), Australia and South Africa to establish smaller independent sugarcane EST programs. The ICSB program to date is that of a single laboratory analyzing three cDNA libraries – apex, mature leaf and mature internode – to develop 9216 ESTs that were clustered into 3400 non-redundant tags (Ma et al., 2004). About 57% of these ESTs were assigned a putative function based on statistically significant similarity to previously characterized proteins or sequences. Another 28% corresponded to previously identified, but uncharacterized, sequences. Some of the remaining sequences were predicted to be genes that may be new or unique to sugarcane. Comparisons of the sugarcane ESTs to a large sorghum EST database revealed similar compositions of expressed genes between different tissues, suggesting applicability of the more abundant sorghum data. Curiously, a substantial fraction of the ICSB ESTs were absent from SUCEST, suggesting that genotype × environment interactions play an important role in the samples of sequences available to molecular biology.

The published sugarcane EST research has profiled gene expression differences between immature stalk internodes that are not storing sucrose versus internodes that are mature and storing sucrose (Carson and Botha, 2002; Carson et al., 2002), or it has focused on internodes that are in the process of maturing and most active in accumulating sucrose (Casu et al., 2003, 2004, 2005). The underlying hypothesis of both approaches is that knowledge about gene expression associated with high storage of sucrose will be revealed through global analysis and can contribute to a systems approach for increasing yield potential. Approximately one-third of the 400 cDNAs analyzed, 200 from each of the two reciprocal subtractive cDNA libraries, were preferentially expressed in either the immature or mature internodes (Carson and Botha, 2002). ESTs generated from all 132 differentially expressed clones revealed 95 unique transcripts of which, based on homology, two-

thirds were assigned functions such as cell wall metabolism, carbohydrate metabolism, stress responses and regulatory proteins. ESTs directly associated with sucrose metabolism were found not to be developmentally regulated, suggesting that growth and maturation of the sugarcane stalk is associated with the expression of genes for a variety of processes other than sucrose metabolism. Likewise, a sequence survey of 7242 ESTs derived from a sucrose-accumulating, maturing stalk revealed that transcripts for carbohydrate metabolism gene sequences (CMGs) were relatively rare in this tissue (Casu et al., 2003). Nevertheless, ESTs of sugar transporter homologues were highly abundant CMG transcripts. The most abundant of the sugar transporter ESTs was associated with phloem companion cells and nearby parenchyma, suggesting a critical role for translocation in sucrose accumulation. The coordinated expression of genes encoding enzymes involved in sucrose synthesis and cleavage as well as glycolysis and the pentose phosphate pathway points toward the need for systems-level approaches to understand sucrose accumulation in sugarcane.

Such systems-level approaches are just beginning (Casu et al., 2004; also see accompanying papers of Casu et al. and Watt et al. in this issue). Scientists in South Africa are focusing on a selected subset of genes, proteins and other small molecules that might be active in sucrose accumulation to develop initial models that will be tested with data from high-throughput technologies including array analyses in comparative studies to detect and quantify the molecular responses to each variable—age, development and genotype, (Watt et al., 2005). Later research will focus on perturbing a single genotype with environmental variables – temperature, water, mineral nutrients – known to effect biochemical changes at the source or sink and then quantify the molecular responses to these perturbations to evolve new hypotheses. Scientists in Australia are focusing on bioinformatic analysis of EST collections from mature and immature stem tissues to suggest functions for genes expressed in the sucrose storing stem and potential interactions among them. They have reported an abundance of several classes of sequence associated with fiber biosynthesis in the maturing stem (Casu et al., 2004a). It is interesting to speculate whether down regulation of some of these genes might redirect photoassimilate from fiber to increased sucrose accumulation. Gene expression profiles, obtained from DNA array analyses, were compared between high sucrose progeny and low sucrose progeny of a segregating population during maturation and sucrose accumulation to develop a 'genetical genomics' strategy for identifying candidate genes that may control sugar accumulation (Casu et al., 2005). The programs of both South Africa and Australia are examples where recent rapid expansion of sugarcane molecular datasets and the beginning of a systems approach to metabolic modeling of sucrose accumulation are pointing toward future applications for raising sucrose yields of sugarcane.

7. Synthesis and prospects

The ultimate goal of biological science is to understand life forms in sufficient detail to allow predictions about how the system will perform when it is placed in a different environment or it is genetically changed. Although this goal has not been achieved, progress is being made rapidly through systems biology approaches applied to prokaryotic organisms. For 43 organisms having relatively complete metabolic information (Jeong et al., 2000), only about 4% of their metabolites were shared but key metabolic pathways and motifs were frequently shared. An even higher level of sharing might be expected at the module level, but data were not sufficient to demonstrate this. It is generally believed that key properties of functional modules will be shared across most species (Oltvai and Barabasi, 2002). If this is true, then knowledge gained from these simpler life forms will provide insights into complex interactions such as those involved in sucrose accumulation in sugarcane.

Successful physiological analysis requires an understanding of the functional interactions between the key components of cells, organs and systems within the living organism. This information resides neither in the genome nor even in the individual proteins encoded by the genes, but rather at the level of protein interactions within the context of subcellular, cellular, tissue, organ and system structures (Noble, 2002). Therefore, it is up to us scientists to develop models that copy nature and to compute these interactions for predicting how each hierarchical level

of complexity might respond to changes. The rapid growth in biological databases and in models of cells, tissues and organs, coupled with the development of powerful computing hardware and algorithms, has made it possible to explore functionality in a quantitative manner all the way from the level of genes to the physiological function of whole organs and regulatory systems. Such approaches have been used recently for modeling complex mutli cellular development and physiology of the human heart (Noble, 2002) and development of the body plan of the sea urchin embryo (Davidson et al., 2002). These recent successes in modeling to increase understanding of complex life forms suggest that the systems approach of top–down modeling coupled with conventional and high-throughput reductionist approaches have great potential for increasing our understanding of sucrose accumulation in sugarcane. To accomplish this goal will require biologists working with experts in bioinformatics and modeling as outlined in Fig. 7. The blueprint is in place and it is time to follow the mandate of Nobel Laureate Sydney Brenner to ''get on with the biology,'' while being fully cognizant of the complexity of the problem as foreseen by another Nobel Laureate, Stephen Hawking, who said, *"I think the next century will be the century of complexity."*

Acknowledgments

I wish to thank Mr. James Carr, USDA, ARS, Pacific Basin Agricultural Research Center for the graphics, Ms Linda Whitehand, USDA, ARS, Western Regional Research Center for assistance with the crop yield analyses, and Dr. Hairong Wei for helpful discussions.

References

Asnaghi, C., Paulet, F., Kaye, C., Grivet, L., Deu, M., Glaszmann, J.C., D'Hont, A., 2000. Application of synteny across Poacee to determine the map location of a sugarcane rust resistance gene. Theor. Appl. Genet. 101, 962–969.

Asnaghi, C., Roques, D., Ruffel, S., Kaye, C., Hoarau, J-Y., Télismart, H., Girard, J.C., Raboin, L.M., Risterucci, A.M., Grivet, L., D'Hont, A., 2004. Targeted mapping of a sugarcane rust resistance gene (*Bru*1) using bulked segregant analysis and AFLP markers. Theor. Appl. Genet. 108, 759–764.

Austin, R.B., 1999. Yield of wheat in the United Kingdom: recent advances and prospects. Crop Sci. 39, 1604–1610.

Barabási, A-L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.

Bull, T.A., Glasziou, K.T., 1963. The evolutionary significance of sugar accumulation in *Saccharum*. Aust. J. Biol. Sci. 16, 737–742.

Carson, D.L., Botha, F.C., 2002. Genes expressed in sugarcane maturing internodal tissue. Plant Cell Rep. 20, 1075–1081.

Carson, D.L., Huckett, B.I., Botha, F.C., 2002. Sugarcane ESTs differentially expressed in immature and maturing internodal tissue. Plant Sci. 162, 289–300.

Casu, R.E., Grof, C.P.L., Rae, A.L., McIntyre, L., Dimmock, C.M., Manners, J.M., 2003. Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Mol. Biol. 52, 371–386.

Casu, R.E., Dimmock, C.M., Chapman, S.C., Grof, C.P.L., McIntyre, L., Bonnett, G.D., Manners, J.M., 2004. Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. Plant Mol. Biol. 54, 503–517.

Casu, R.E., Manners, J.M., Bonnett, G.D., Jackson, P.A., McIntyre, L., Dunne, R., Chapman, S.C., Rae, A.L., Grof, C.P.L., 2005. Genomics approaches for the identification of genes determining important traits in sugarcane. Field Crops Res. 92, 137–147.

Cox, M.V., Hogarth, D.M., Smith, G.R., 2000. Cane breeding and improvement. In: Hogarth, D.M., Alsopp, P.G. (Eds.), Manual of Cane Growing. Bureau of Sugar Experiment Stations, Brisbane, Australia, pp. 91–108.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Titus Brown, C., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z.J., Schilstra, M.J., Clarke, P.J.C., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002. A genomic regulatory network for development. Science 295, 1669–1678.

Duvick, D.N., 1996. Plant Breeding, an evolutionary concept. Crop Sci. 36, 539–548.

Duvick, D.N., Cassman, K.G., 1999. Post-green revolution trends in yield potential of temperate maize in the north-central United States. Crop Sci. 39, 1622–1630.

Evans, L.T., 1975. Crops and world food supply, crop evolution, and the origins of crop physiology. In: Evans, L.T. (Ed.), Crop Physiology, Some Case Histories. Cambridge University Press, London, pp. 1–22.

Evans, L.T., Fischer, R.A., 1999. Yield potential: its definition, measurement, and significance. Crop Sci. 39, 1544–1551.

Ideker, T., Thorsson, V., Ranish, J.A., Christama, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebrsold, R., Hood, L., 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292, 929–934.

Irvine, J.E., 1975. Relations of photosynthetic rates and leaf and canopy characters to sugarcane yield. Crop Sci. 15, 671–676.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.-L., 2000. The large-scale organization of metabolic networks. Nature 407, 651–654.

Kitano, H., 2002. Systems biology: a brief overview. Science 295, 1662–1664.

Le Cunff, L., Garsmeur, O., Raboin, L.M., Grivet, L., Glaszmann, J.C., D'Hont, A., 2004. Construction of a sugarcane BAC contig covering the target resistance gene (*bru*1) through exploitation of synteny with sorghum and rice. In: Plant and animal genome conference XII abstract.

Loomis, R.S., Amthor, J.S., 1999. Yield potential, plant assimilatory capacity and metabolic efficiencies. Crop Sci. 39, 1584–1596.

Lunn, J.E., MacRae, E., 2003. New complexities in the synthesis of sucrose. Curr. Opin. Plant Biol. 6, 208–214.

Ma, H.-M., Schulze, S., Lee, S., Yang, M., Mirkov, E., Irvine, J., Moore, P., Paterson, A., 2004. An EST survey of the sugarcane transcriptome. Theor. Appl. Genet. 108, 851–863.

Meredith, W.R., Bridge, R.R., 1984. Genetic contributions to yield gains in upland cotton. In: Fehr, W.R. (Ed.), Genetic Contributions to Yield Gains of Five Major Crop Plants. Crop Science Society of America Special Publication Number 7, pp. 75–87.

Miller, F.R., Kebede, Y., 1984. Genetic contributions to yield gains in sorghum, 1950–1980. In: Fehr, W.R. (Ed.), Genetic Contributions to Yield Gains of Five Major Crop Plants. Crop Science Society of America Special Publication Number 7, pp. 1–14.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alton, U., 2002. Network motifs: simple building blocks of complex networks. Science 298, 824–827.

Ming, R., Liu, S.C., Moore, P.H., Irvine, J.E., Paterson, A.H., 2001. QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res. 11, 2075–2084.

Ming, R., DelMonte, T., Moore, P.H., Irvine, J.E., Paterson, A.H., 2002. Comparative analysis of QTLs affecting plant height and flowering time among closely-related diploid and polyploid genomes. Genome 45, 794–803.

Minorsky, P.V., 2003. Achieving the *in silico* plant. Systems biology and the future of plant biological research. Plant Physiol. 132, 404–409.

Noble, D., 2002. Modeling the heart—from genes to cells to the whole organ. Science 295, 1678–1682.

Oltvai, A.N., Barabási, A.-L., 2002. Life's complexity pyramid. Science 298, 763–764.

Overbeek, R., Larsen, N., Pusch, G.D., D'Souza, M., Selkov, E., Kyrpides, N., Fonstein, M., Maltsev, N., Selkov, E., 2000. WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. Nucleic Acids Res. 28, 123–125.

Peng, S., Cassman, K.G., Virmani, S.S., Sheehy, J., Khush, G.S., 1999. Yield potential trends of tropical rice since the release of IR8 and the challenge of increasing rice yield potential. Crop Sci. 39, 1552–1559.

Pingali, P.L., Moya, P.F., Velasco, L.E., 1990. The post-green revolution blues in Asian rice production: the diminished gap between experiment station and farmer yields. IRRI Social Science Paper 90-01. International Rice Research Institute (IRRI), Manila, Philippines.

Rabbinage, R., 1993. The ecological background of food production. In: Chadwick, D.J., Marsh, J. (Eds.), Crop Protection and Sustainable Agriculture. Ciba Foundation Symposium, vol. 177. John & Wiley and Sons, Chicester, pp. 2–29.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.-L., 2002. Hierarchical organization of modularity in metabolic networks. Science 297, 1551–1555.

Reynolds, M.P., Rajaram, S., Sayre, K.D., 1999. Physiological and genetic changes of irrigated wheat in the post-green revolution period and approaches for meeting projected global demand. Crop Sci. 39, 1611–1621.

Sinclair, T.R., Muchow, R.C., 1999. Radiation use efficiency. Adv. Agron. 65, 215–265.

Specht, J.E., Hume, D.J., Kumudini, S.V., 1999. Soybean yield potential—a genetic and physiological perspective. Crop Sci. 39, 1560–1570.

Tollenaar, M., Wu, J., 1999. Yield improvement in temperate maize is attributable to greater stress tolerance. Crop Sci. 39, 1597–1604.

Vettore, A.L., da Silva, F.R., Kemper, E.L., Arruda, P., 2001. The libraries that made SUCEST. Gen. Mol. Biol. 24 (14), 1–7.

Vettore, A.L., da Silva, F.R., Kemper, E.L., Souza, G.M., da Silva, A.M., Ferro, M.I.T., Henrique-Silva, F., Giglioti, E.A., Lemos, M.V.F., Coutinho, L.L., Nobrega, M.P., Carrer, H., Franca, S.C., Bacci Jr., M., Goldman, M.H.S., Gomes, S.L., Nunes, L.R., Camarfgo, L.E.A., Siqueira, W.J., Van Sluys, M-A., Thiemann, O.H., Kuramae, E.E., Santelli, R.V., Marino, C.L., Targon, M.L.P.N., Ferro, J.A., Silveira, H.C.S., Marini, D.C., Lemos, E.G.M., Monteiro-Vitorello, C.B., Tambor, J.H.M., Carraro, D.M., Roberto, P.G., Martins, V.G., Goldman, G.H., de Oliveira, G.C., Truffi, D., Colombo, C.A., Rossi, M., de Araujo, P.G., Sculaccio, S.A., Angella, A., Lima, M.M.A., Grivet, L., Di Mauro, S.M.Z., Nobrega, F.G., Menck, C.F.M., Braga, M.D.V., Telles, G.P., Cara, F.A.A., Pendrosa, G., Meidanis, J., Arruda, P., 2003. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res. 13, 2725–2735.

Watson, J.D., Crick, F.H.C., 1953. A structure for deoxyribose nucleic acid. Nature 171, 737.

Watt, D. McCormick, A., Govender, C., Cramer, M., Huckett, B. Botha, F., 2005. Increasing the utility of genomics in unraveling sucrose accumulation. Field Crops Res., this issue.

Wu, K.K., Burnquist, W., Sorrells, M.E., Tew, T.L., Moore, P.H., Tanksley, S.D., 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. 83, 294–300.